ELSEVIER

Computational Statistics & Data Analysis 32 (1999) 29–46

COMPUTATIONAL STATISTICS & DATA ANALYSIS

www.elsevier.com/locate/csda

A unified approach to estimating population size for a births only model

Cathy W.S. Chen[a,*], Shen-Ming Lee[a], Ying-Hen Hsieh[b], Kumnuan Ungchusak[c]

[a] *Department of Statistics, Feng-Chia University, Taichung 407, Taiwan*
[b] *Department of Applied Mathematics, National Chung-Hsing University, Taichung, Taiwan*
[c] *Division of Epidemiology, Thailand Ministry of Public Health, Bangkok, Thailand*

Received 1 August 1998; received in revised form 1 December 1998; accepted 18 March 1999

Abstract

The primary goal of this paper is to estimate the population size for a births only model when the capture probabilities vary with behavior response and time (or sampling occasion). A Bayesian framework is developed from the births only models for the capture–recapture experiment. We propose a unified approach for estimating the population size on each sampling occasion for four specified models using the Gibbs sampler, a Markov chain Monte Carlo method. The proposed methodology is illustrated with a simulation study and HIV serosurveillance data of Thailand. The results show that Gibbs sampler provides a reasonable estimate of population size when compared with the classical technique.

Keywords: Capture–recapture sampling; Gibbs sampler; Markov chain Monte Carlo method; Behavior response; Time variation; Births only model; HIV serosurveillance data; HIV-infected IVDU's; Open population

1. Introduction

In ecological studies, the first problem usually encountered is the need to know the size of various wildlife populations in our study area, and to know how these populations change with time. In order to make more precise inferences various capture–recapture sampling techniques have been widely used. Seber (1982) classified

* Corresponding author. Fax: +886-886-4451-7092.
E-mail address: chenws@fcu.edu.tw (C.W.S. Chen)

0167-9473/99/$ - see front matter
PII: S 0167-9473(99)00023-7

population into two categories called "closed" and "open", depending on whether the population remains unchanged during the period of investigation, or changes through such processes as birth, mortality, emigration, etc. Like the closed population models, open models can vary widely in their generality, and as a result many models have been introduced into the literature by Jolly (1965), Seber (1982), Cormack (1968, 1972), etc. The births only model proposed by Darroch (1959) and Jolly (1965) investigates the population size when additions by birth, recruitment, or immigration is possible. In this work, we focus mainly on the births only model which allows only births to occur during the experiment. We will consider the problem of estimating population size for all sampling periods for a births only model when capture probabilities vary with behavior response and time (or sampling occasion). This model is called $M_{tb}^{(B)}$, where B denotes the births only model, t denotes time variation, and b denotes behavior response. There are three special cases of Model $M_{tb}^{(B)}$: Model $M_t^{(B)}$, Model $M_b^{(B)}$, and Model $M_0^{(B)}$. Model $M_t^{(B)}$ and Model $M_b^{(B)}$ consider the capture probabilities varying with time and behavior response, respectively. Model $M_0^{(B)}$ denotes that the capture probability remains constant between all sampling periods for a births only model. For related work in this area, see Pollock et al. (1990), Seber (1986, 1992) and Hwang and Chao (1995).

As far as we know, there is no work on Model $M_b^{(B)}$ and Model $M_{tb}^{(B)}$ in the literature, only Model $M_t^{(B)}$ has been studied by Darroch (1959) and Jolly (1965) to estimate the population size for each sampling period. However, the population size of the first period is unestimable. George and Robert (1992) and Lee and Chen (1998) proposed Bayesian inferences about the population size for various models on a closed population. In this paper, we generalize the Bayesian analysis of the above-mentioned articles to the births only model using the Gibbs sampler, a Markov chain Monte Carlo method. We shall not repeat the details of the Gibbs sampler which can be found elsewhere (e.g. Geman and Geman, 1994, Tanner and Wong, 1987; Gelfand and Smith, 1990; Gelfand et al., 1990). It suffices to say that what we need are conditional distributions of subsets of parameters given the others. The Gibbs sampler is iterated many times in order to obtain a sample of draws from the posterior distribution. The empirical distribution of this sample converges weakly to the true joint distribution. For more details of the convergence results, see Tierney (1994). Interested readers are also referred to Casella and George (1992) and Tanner (1994) for a general comprehensive review of the Gibbs sampler. Section 2 presents the births only model and the setup of a Bayesian framework. In Section 3 we illustrate the methodology using a simulation study and the semi-annual serosurveillance data of Thailand. We give concluding remarks in Section 4.

2. Inference for the births only model

We consider a sequence of s samples. Let t_j be the time when the jth sample is taken and let B_j be the number of new animals or subjects entering the population between time t_j and time t_{j+1}. Assume that the B_k animals are indexed by $1,\ldots,B_k$, for $k=0,\ldots,s-1$, and $P_{ij}^{(k)}$ is the capture probability of the ith animal entering the

population between time t_k and t_{k+1} in sample j, with $i=1,\ldots,B_k$, $k=0,\ldots,s-1$, and $j=k+1,\ldots,s$. We define N_k be the total number animals in the population just before time t_k, and $N_k = B_0 \cdots + B_{k-1}$ in the births only model.

In this article, animals are assumed to act independently. When there exists a behavior response effect, $P_{ij}^{(k)}$ will depend on the capture history of this animal from the $(k+1)$th to the $(j-1)$th sample, and $P_{ij}^{(k)}$ can be expressed as

$$P_{ij}^{(k)} = \begin{cases} P_{ij}^{(k)'} & \text{if the } i\text{th animal of } B_k \text{ is not caught between the } (k+1)\text{th} \\ & \text{to the } (j-1)\text{th sample;} \\ b_{ij}^{(k)'} & \text{if the } i\text{th animal of } B_k \text{ has been caught between the } (k+1)\text{th} \\ & \text{to the } (j-1)\text{th sample.} \end{cases} \tag{1}$$

Let $X_{ij}^{(k)}$ be equal to 1 if the ith animal entering the population between time t_k and t_{k+1} is caught in the jth sample, and equal to 0 otherwise. The capture sequence of the ith animal entering the population between time t_k and t_{k+1} is $\{X_{ij}^{(k)}, j=k+1,\ldots,s\}$. From (1), the probability function of $X_{ij}^{(k)}$ depends on the previous capture history of the ith animal. Hence the underlying general probability structure of the capture–recapture experiments for a births only model is

$$\begin{aligned} L(\boldsymbol{B},\boldsymbol{P}\mid\mathscr{D}) &= \prod_{k=0}^{s-1}\prod_{i=1}^{B_k}\prod_{j=k+1}^{s}(P_{ij}^{(k)})^{X_{ij}^{(k)}}(1-P_{ij}^{(k)})^{1-X_{ij}^{(k)}} \\ &= \prod_{k=0}^{s-1}\prod_{i=1}^{B_k}\prod_{j=k+1}^{s}(P_{ij}^{(k)'})^{X_{ij}^{(k)}I[(\sum_{r=k+1}^{j-1}X_{ir}^{(k)})=0]}(b_{ij}^{(k)'})^{X_{ij}^{(k)}I[(\sum_{r=k+1}^{j-1}X_{ir}^{(k)})>0]} \\ &\quad\times(1-P_{ij}^{(k)'})^{(1-X_{ij}^{(k)})I[(\sum_{r=k+1}^{j-1}X_{ir}^{(k)})=0]}(1-b_{ij}^{(k)'})^{(1-X_{ij}^{(k)})I[(\sum_{r=k+1}^{j-1}X_{ir}^{(k)})>0]}, \end{aligned} \tag{2}$$

where $I(.)$ is the usual indicator function, $\boldsymbol{B}=(B_0,\ldots,B_{s-1})$, $\boldsymbol{P}=(P_{ij}^{(k)},\ i=1,\ldots,B_k;\ j=k+1,\ldots,s;k=0,\ldots,s-1)$, and $\mathscr{D}=\{X_{ij}^{(k)},\ i=1,\ldots,B_k;\ j=k+1,\ldots,s;\ k=0,\ldots,s-1\}$. There are too many parameters in the general model in (2), so the information about B_k or N_k cannot be extracted from data. Therefore, the parameter space of the general model in (2) must be restricted. The most common restrictions used are $P_{ij}^{(k)}=P$ or $P_{ij}^{(k)}=P_j$ (see Jolly, 1965 or Seber, 1982) for $i=1,\ldots,B_k$, and $k=0,\ldots,j-1$. The models are designated Model $M_0^{(B)}$ or $M_t^{(B)}$, respectively. If there is behavior response for the captured animal, the respective restrictions become

$$P_{ij}^{(k)} = P_j I\left(\sum_{r=k+1}^{j-1}X_{ir}^{(k)}=0\right)+b_j I\left(\sum_{r=k+1}^{j-1}X_{ir}^{(k)}>0\right) \tag{3}$$

or

$$P_{ij}^{(k)} = PI\left(\sum_{r=k+1}^{j-1}X_{ir}^{(k)}=0\right)+bI\left(\sum_{r=k+1}^{j-1}X_{ik}^{(k)}>0\right), \tag{4}$$

where b_j is the recapture probability in jth sample and b is the recapture probability for any sample. We denote (3) and (4) by Model $M_{tb}^{(B)}$ and Model $M_b^{(B)}$, respectively. In this paper, we will give a unified inference of population size for each sampling period via Gibbs sampler on the above four Models: $M_{tb}^{(B)}$, $M_b^{(B)}$ $M_t^{(B)}$, and $M_0^{(B)}$.

2.1. Model $M_{tb}^{(B)}$

In this subsection, we consider the case when there are behavior response and time variation for the capture probability. We assume that all animals in population just before time t_j which have not been caught in the first $j-1$ samples have the same capture probabilities P_j in the jth sample, and that the recapture probabilities for all animals ever captured in population just before time t_j in the jth sample are b_j. The explicit structure of $P_{ij}^{(k)}$ is given in (3). Let $\mathscr{D} = \{u_1, \ldots, u_s, m_2, \ldots, m_s\}$, the likelihood function can be obtained as a special case of (2).

$$L(\boldsymbol{N}, \boldsymbol{P}, \boldsymbol{b} \mid \mathscr{D}) = \left\{ \prod_{j=1}^{s} \binom{N_j - M_j}{u_j} P_j^{u_j} (1-P_j)^{N_j - M_{j+1}} \right\} \prod_{j=2}^{s} \left\{ \binom{M_j}{m_j} b_j^{m_j} (1-b_j)^{M_j - m_j} \right\}, \tag{5}$$

where $\boldsymbol{N} = (N_1, \ldots, N_s)$, $N_j = B_0 + \ldots + B_{j-1}$, $\boldsymbol{P} = (P_1, \ldots, P_s)$, $\boldsymbol{b} = (b_2, \ldots, b_s)$, $M_{j+1} = u_1 + \cdots + u_j$ is the number of distinct animals captured in the first j samples, and u_j and m_j are the number of unmarked and marked animals captured in the jth sample, respectively. Also $m_1 = 0$.

The likelihood function can be expressed as the product of some binomial distributions. The conditional distributions of m_j given M_j and u_j given $N_j - M_j$ are, respectively,

$$m_j \mid M_j \sim B(b_j, M_j), \quad j = 2, \ldots, s$$

and

$$u_j \mid N_j - M_j \sim B(P_j, N_j - M_j), \quad j = 1, \ldots, s.$$

Suppose the prior distribution of $(\boldsymbol{N}, \boldsymbol{P}, \boldsymbol{b})$ is chosen to be

$$\pi(\boldsymbol{N}, \boldsymbol{P}, \boldsymbol{b}) = \left(\prod_{j=1}^{s} \pi(P_j) \right) \left(\prod_{j=2}^{s} \pi(b_j) \right) \pi(N_1, \ldots, N_s).$$

In addition, let $\pi(P_j) = Be(\gamma_1, \gamma_2)$ and $\pi(b_j) = Be(\gamma_3, \gamma_4)$, with $Be(x, y)$ denoting a beta distribution. It follows that the complete conditional posterior distributions are given by

$$\pi(\boldsymbol{N} \mid \boldsymbol{P}, \boldsymbol{b}, \mathscr{D}) \propto \prod_{j=1}^{s} \binom{N_j - M_j}{u_j} (1-P_j)^{N_j} \pi(N_1, \ldots, N_s), \tag{6}$$

$$\pi(\boldsymbol{P}|\boldsymbol{N},\boldsymbol{b},\mathscr{D}) = \prod_{j=1}^{s} Be(u_j+\gamma_1, N_j - M_{j+1} + \gamma_2), \tag{7}$$

$$\pi(\boldsymbol{b}|\boldsymbol{N},\boldsymbol{P},\mathscr{D}) = \prod_{j=2}^{s} Be(m_j+\gamma_3, M_j - m_j + \gamma_4). \tag{8}$$

Note that the conditional posterior of $\boldsymbol{N}$ given $(\boldsymbol{P},\boldsymbol{b},\mathscr{D})$ does not depend on the recapture probabilities $\boldsymbol{b}$, therefore the information of $\boldsymbol{N}$ only depends on $\boldsymbol{P}$. When $B_k = 0$, $k = 1,\ldots,s-1$, that is, there are no new animals entering the population between time t_2 to time t_s, the births only model degenerates to a closed population model. In this case, we have $N_1 = \cdots = N_s$, and the conditional posteriors (6)–(7) become the same as those proposed by Lee and Chen (1998).

Let $\boldsymbol{N}_{(-j)}$ denote the vector $\boldsymbol{N}$ with the N_j deleted. If the prior of $\boldsymbol{N}$ is constant, then the conditional posterior probability function of N_j is

$$\begin{aligned}
&\pi(N_j \mid \boldsymbol{N}_{(-j)},\boldsymbol{P},\boldsymbol{b},\mathscr{D}) \\
&= \frac{\left\{\prod_{l=1,l\neq j}^{s}\binom{N_l-M_l}{u_l} P_l^{u_l}(1-P_l)^{N_l-M_{l+1}}\right\}\binom{N_j-M_j}{u_j} P_j^{u_j}(1-P_j)^{N_j-M_{j+1}}}{\sum_{N_j=\max\{N_{j-1},M_{j+1}\}}^{N_{j+1}}\left\{\prod_{l=1,l\neq j}^{s}\binom{N_l-M_l}{u_l} P_l^{u_l}(1-P_l)^{N_l-M_{l+1}}\right\}\binom{N_j-M_j}{u_j} P_j^{u_j}(1-P_j)^{N_j-M_{j+1}}} \\
&= \frac{\binom{N_j-M_j}{u_j} P_j^{u_j}(1-P_j)^{N_j-M_{j+1}}}{\sum_{N_j=\max\{N_{j-1},M_{j+1}\}}^{N_{j+1}}\binom{N_j-M_j}{u_j} P_j^{u_j}(1-P_j)^{N_j-M_{j+1}}}.
\end{aligned} \tag{9}$$

Note that $\pi(N_j \mid \boldsymbol{N}_{(-j)},\boldsymbol{P},\boldsymbol{b},\mathscr{D})$ is a truncated negative binomial. Since $M_{j+1} \le N_j$ and $N_{j-1} \le N_j \le N_{j+1}$, the right truncated and left truncated points are N_{j+1} and $\max\{N_{j-1}, M_{j+1}\}$, respectively. Starting with an initial value of $\{N_1^{(0)},\ldots,N_s^{(0)}\}$ and given fixed $(\gamma_1,\gamma_2,\gamma_3,\gamma_4)$, we can produce a 'Gibbs sequence' $\{\boldsymbol{P}^{(k)},\boldsymbol{N}^{(k)},\boldsymbol{b}^{(k)}\}$, $k = 0,\ldots$ with simulated sampling from (7)–(9). Note that the initial value of $\{N_1^{(0)},\ldots, N_s^{(0)}\}$ has to satisfy $N_1^{(0)} \le \cdots \le N_s^{(0)}$ and the inference for $\boldsymbol{N}$ does not depend on the hyper-parameters (γ_3,γ_4).

2.2. Model $M_b^{(B)}$

For Model $M_b^{(B)}$, the capture probabilities of all animals may vary with behavior response. Therefore, all animals in the population just before each capture time t_j, $j = 1,\ldots,s$, have the same capture probabilities P in the first capture, and the same recapture probabilities b after the first capture. Subsequently, the structure capture probabilities $P_{ij}^{(k)}$ in (1) can reduce to (4) and the likelihood function in (5) becomes

$$\begin{aligned}
L(\boldsymbol{N},P,b \mid \mathscr{D}) = &\left\{\prod_{j=1}^{s}\binom{N_j-M_j}{u_j}\right\} P^{M_{s+1}}(1-P)^{N_\bullet - M_\bullet - M_{s+1}} \\
&\times \left\{\prod_{j=2}^{s}\binom{M_j}{m_j}\right\} b^{m_\bullet}(1-b)^{M_\bullet - m_\bullet},
\end{aligned} \tag{10}$$

where $N_\bullet = N_1 + \cdots + N_s$, $M_\bullet = M_2 + \cdots + M_s$ and $m_\bullet = m_2 + \cdots + m_s$. Suppose that the prior distribution of $(\boldsymbol{N}, P, b)$ is given by $\pi(\boldsymbol{N}, P, b) = \pi(N_1, \ldots, N_s)\pi(P)\pi(b)$, which asserts that $\boldsymbol{N}$, P, and b are priori independent. In addition, let $\pi(P) = Be(\gamma_1, \gamma_2)$ and $\pi(b) = Be(\gamma_3, \gamma_4)$. Such priors lead to conditional posteriors of the forms:

$$\pi(\boldsymbol{N} \mid P, b, \mathscr{D}) \propto \left\{ \prod_{j=1}^{s} \binom{N_j - M_j}{u_j} \right\} (1-P)^{N_\bullet} \pi(N_1, \ldots, N_s), \tag{11}$$

$$\pi(P \mid \boldsymbol{N}, b, \mathscr{D}) = Be(\gamma_1 + M_{s+1}, \gamma_2 + N_\bullet - M_\bullet - M_{s+1}), \tag{12}$$

$$\pi(b \mid \boldsymbol{N}, P, \mathscr{D}) = Be(\gamma_3 + m_\bullet, \gamma_4 + M_\bullet - m_\bullet). \tag{13}$$

Other prior distributions can also be used. For example, if $\pi(N_1, \ldots, N_s)$ is constant, then the conditional posterior of N_j given $(\boldsymbol{N}_{(-j)}, P, b, \mathscr{D})$ follows a truncated negative binomial. Moreover, the left and right truncated points are $\max\{N_{j-1}, M_{j+1}\}$ and N_{j+1}. The explicit conditional posterior of N_j is

$$\pi(N_j \mid \boldsymbol{N}_{(-j)}, P, b, \mathscr{D}) = \frac{\binom{N_j - M_j}{u_j} P^{u_j} (1-P)^{N_j - M_{j+1}}}{\sum_{N_j = \max\{N_{j-1}, M_{j+1}\}}^{N_{j+1}} \binom{N_j - M_j}{u_j} P^{u_j} (1-P)^{N_j - M_{j+1}}}. \tag{14}$$

Note that the above equation is the same as (9), but we keep $P_j = P$ on each capture sampling time t_j. Now choose $(\gamma_1, \gamma_2, \gamma_3, \gamma_4)$ and let the initial value of $\boldsymbol{N}$ be $\boldsymbol{N}^{(0)} = (N_1^{(0)}, \ldots, N_s^{(0)})$ satisfying $N_1^{(0)} \leq \cdots \leq N_s^{(0)}$, the Gibbs sequence $\{\boldsymbol{N}^{(k)}, P^{(k)}, b^{(k)}\}$, $k = 1, \ldots$ can be obtained from (12)–(14).

2.3. Model $M_t^{(B)}$

Suppose all animals in the population just before time t_j have the same capture probability P_j which only depends on the capture sampling time t_j, $j = 1, \ldots, s$. This model, Model $M_t^{(B)}$, has been studied by Darroch (1959) and Jolly (1965). The likelihood function of Model $M_t^{(B)}$ can be reduced to

$$L(\boldsymbol{N}, \boldsymbol{P} \mid \mathscr{D}) = \left\{ \prod_{j=1}^{s} \binom{N_j - M_j}{u_j} P_j^{n_j} (1-P_j)^{N_j - n_j} \right\} \left\{ \prod_{j=2}^{s} \binom{M_j}{m_j} \right\}, \tag{15}$$

where $n_j = u_j + m_j$, $j = 1, \ldots, s$, are the total number of captured animals in sample j. We choose the priors of $(\boldsymbol{N}, \boldsymbol{P})$ as $\pi(\boldsymbol{N}, \boldsymbol{P}) = \pi(N_1, \ldots, N_s) \prod_{j=1}^{s} \pi(P_j)$ and $\pi(P_j) = Be(\gamma_1, \gamma_2)$ for $j = 1, \ldots, s$. As a result, the conditional posterior of $\boldsymbol{N}$ is the same as Eq. (6) for Model $M_{tb}^{(B)}$. Moreover, the conditional posterior of $\boldsymbol{P}$ given $(\boldsymbol{N}, \mathscr{D})$ can be reduced to

$$\pi(\boldsymbol{P} \mid \boldsymbol{N}, \mathscr{D}) = \prod_{j=1}^{s} Be(\gamma_1 + n_j, \gamma_2 + N_j - n_j).$$

Therefore, given $(N_1, \ldots, N_s)$ and $\mathscr{D}$, we can simulate $\boldsymbol{P}$ using the above result. When we choose $\pi(N_1, \ldots, N_s)$ to be constant, the conditional posterior distribution

of N_j given $(\boldsymbol{N}_{(-j)}, \boldsymbol{P}, \mathscr{D})$ also is a truncated negative binomial and the form of this conditional posterior is the same as in (9). Moreover, we can simulate the Gibbs sequence $\{\boldsymbol{N}^{(k)}, \boldsymbol{P}^{(k)}\}$, $k=1,\ldots$ for the given initial value $\boldsymbol{N}^{(0)}$ which have the same constraints as in Model $M_{tb}^{(B)}$ and $M_b^{(B)}$.

2.4. Model $M_0^{(B)}$

In this subsection, we consider $M_0^{(B)}$, a special case of $M_t^{(B)}$. Suppose all animals in the population just before each sampling time t_j, $j=1,\ldots,s$, have the same capture probabilities P. The capture probabilities $P_{ij}^{(B)}$ is constant for $k=0,\ldots,s-1$, $j=k+1,\ldots,s$, and $i=1,\ldots,B_k$. The likelihood function in Eq. (15) can be rewritten as

$$L(\boldsymbol{N}, P \mid \mathscr{D}) = \left\{\prod_{j=1}^{s} \binom{N_j - M_j}{u_j}\right\} \left\{\prod_{j=2}^{s} \binom{M_j}{m_j}\right\} P^{n_\bullet}(1-P)^{N_\bullet - n_\bullet},$$

where $n_\bullet = n_1 + \cdots + n_s$ and $N_\bullet$ is as defined in Section 2.2. When the priors $(\boldsymbol{N}, P)$ are chosen as $\pi(\boldsymbol{N}, P) = \pi(N_1,\ldots,N_s)\pi(P)$, the conditional posteriors are

$$\pi(\boldsymbol{N} \mid P, \mathscr{D}) = \prod_{j=1}^{s} \binom{N_j - M_j}{u_j} (1-P)^{N_\bullet} \pi(N_1,\ldots,N_s)$$

and

$$\pi(P \mid \boldsymbol{N}, \mathscr{D}) \propto \prod_{j=1}^{s} P^{n_\bullet}(1-P)^{(N_\bullet - n_\bullet)} \pi(P).$$

The above conditional posterior distribution of $\boldsymbol{N}$ given $(P, \mathscr{D})$ is the same as the conditional posterior of $\boldsymbol{N}$ for Model $M_b^{(B)}$, while the conditional posterior distribution of P is different for Model $M_b^{(B)}$ and Model $M_0^{(B)}$. The above result depends on the total number of captured animals from all samplings. Let $\pi(P) = Be(\gamma_1, \gamma_2)$, the conditional posterior, then $\pi(P \mid \boldsymbol{N}, \mathscr{D})$ becomes

$$\pi(P \mid \boldsymbol{N}, \mathscr{D}) = Be(\gamma_1 + n_\bullet, \gamma_2 + N_\bullet - n_\bullet).$$

Therefore, P can be generated easily. Moreover, if we choose a constant prior for $\boldsymbol{N}$, the conditional posterior distribution of N_j can be obtained which is the same as the conditional posterior of N_j in Model $M_b^{(B)}$.

3. Illustrative examples

3.1. Simulation study

In this section, we carried out a simulation study to investigate the performance of the proposed inference procedure. We choose the number of new animals entering population as $(B_0,\ldots,B_{s-1})=(300,100,100,100,100)$ where $s=5$. Let the first capture probability of animals of the jth sample be P_j and let δ be the behavior response factor. We define the recapture probability $b_j = \delta P_j$, where $\delta > 1$ denotes trap-happy,

$\delta < 1$ denotes trap-shy, and $\delta = 1$ if there is no behavior response effect. We consider the following combinations:

1. Three levels of δ: 0.8, 1, and 1.5.
2. Two levels of $(P_1, \ldots, P_5)$: (1) (0.2,0.2,0.2,0.2,0.2) (2) (0.3,0.2,0.1,0.15,0.25).
3. Four choices of prior distributions: (1) *Beta*(0.5,0.5), (2) *Beta*(10,40), (3) *Beta(5,45)*, (4) *Beta*(15,35).

We apply the Bayesian approach to our simulation. For each data set, we develop inference for $(N_1, \ldots, N_5)$ via the proposed method for Models $M_0^{(B)}$, $M_t^{(B)}$, $M_b^{(B)}$, and $M_{tb}^{(B)}$, respectively. For comparison, we calculate the maximum likelihood estimates of N_j, $j = 2, \ldots, s$ for Model $M_t^{(B)}$. Note that the ML estimators of population size for Models $M_0^{(B)}$, $M_b^{(B)}$, and $M_{tb}^{(B)}$ do not have explicit forms in the literature. The estimators proposed by Jolly (1965) have the closed form

$$\hat{N}_j = \frac{n_j M_j}{m_j}, \quad j = 2, \ldots, s.$$

We refer to this method as "Jolly's method" hereinafter. The hyper-parameters used are $(\gamma_3, \gamma_4) = (3.0, 3.0)$ and 4 specifications for (γ_1, γ_2). The choice $(\gamma_1, \gamma_2) = (0.5, 0.5)$ can be motivated as a noninformative prior.

For each data set, the Gibbs sampler is run for 2500 iterations but we use every 5th of the last 1500 iterations for making inference. One hundred data sets are generated for each combination and analyzed for each of the four models and Jolly's method. For each data set, we calculate the median, mean, standard deviation, and the 95% Bayesian intervals. Due to limited space, the results for $(\gamma_1, \gamma_2) = (5, 45), (15, 35)$ are omitted. Tables 1–12 list the averages of median, mean, and the standard deviation of 100 trials, the actual coverage probabilities of the 95% Bayesian intervals (the fraction of these intervals that contain the true parameter values), and the coverage probabilities for Jolly's method. A 95% confidence interval on the N_j for Jolly's method is

$$(\hat{N}_j - Z_{0.025}\, se(\hat{N}_j), \quad \hat{N}_j + Z_{0.025}\, se(\hat{N}_j)), \tag{16}$$

where $se(\hat{N}_j) = (\hat{N}_j - n_j)(\hat{N}_j - M_j)/m_j$, and $Z_{0.025}$ denotes the upper 2.5 percentage point of the standard normal distribution. Therefore, the coverage probabilities for Jolly's method are the fraction of intervals in (16) that contain the true parameter values. The true models for all tables are listed as follows:

True model	$M_0^{(B)}$	$M_b^{(B)}$		$M_t^{(B)}$	$M_{tb}^{(B)}$	
		Trap-happy $\delta > 1$	Tray-shy $\delta < 1$		Trap-happy $\delta > 1$	Tray-shy $\delta < 1$
Tables	1–2	3–4	5–6	7–8	9–10	11–12

The simulation results show that when the true models are either Model $M_0^{(B)}$ or Model $M_t^{(B)}$, the estimates for Model $M_t^{(B)}$ with noninformative priors have the smallest biases among all other estimates. Moreover, all of the coverage probabilities for Model $M_t^{(B)}$ are above 80%. While N_1 is not estimable using Jolly's (1965) method, the estimates of N_1 for $M_0^{(B)}$ or $M_t^{(B)}$ perform the best in the sense of less

Table 1
Simulation results for 100 runs and the true model is $M_0^{(B)}$; $(P_1, P_2, P_3, P_4, P_5) = (0.2, 0.2, 0.2, 0.2, 0.2)$, the prior is *Beta*(0.5,0.5)

$\delta = 1$		$M_0^{(B)}$	$M_t^{(B)}$	$M_b^{(B)}$	$M_{tb}^{(B)}$	Jolly
$N_1 = 300$	Med	309.9	233.2	873.7	348.9	–
	Mean(Std)	310.9(24.0)	233.0(93.5)	861.2(138.8)	386.7(213.7)	–
	Coverage	0.740	0.950	0.030	1.000	–
$N_2 = 400$	Med	418.0	401.5	1075.6	676.9	–
	Mean(Std)	419.0(26.2)	402.1(47.3)	1060.5(157.3)	694.7(249.5)	399.7(98.7)
	Coverage	0.670	0.900	0.020	0.890	0.860
$N_3 = 500$	Med	518.0	513.7	1247.1	994.0	–
	Mean(Std)	519.2(28.1)	514.2(39.4)	1227.4(170.0)	987.8(255.6)	500.9(77.7)
	Coverage	0.670	0.890	0.020	0.490	0.920
$N_4 = 600$	Med	613.3	613.0	1403.2	1308.0	–
	Mean(Std)	614.3(29.9)	614.1(38.3)	1376.8(179.2)	1278.7(239.5)	589.2(67.5)
	Coverage	0.800	0.830	0.020	0.090	0.920
$N_5 = 700$	Med	722.3	735.3	1586.6	1620.6	–
	Mean(Std)	723.6(33.6)	739.4(48.1)	1547.8(181.8)	1570.6(187.3)	698.5(65.1)
	Coverage	0.800	0.820	0.030	0.030	0.960

Table 2
Simulation results for 100 runs and the true model is $M_0^{(B)}$; $(P_1, P_2, P_3, P_4, P_5) = (0.2, 0.2, 0.2, 0.2, 0.2)$, the prior is *Beta*(10,40)

$\delta = 1.$		$M_0^{(B)}$	$M_t^{(B)}$	$M_b^{(B)}$	$M_{tb}^{(B)}$	Jolly
$N_1 = 300$	Med	309.1	301.7	416.8	305.4	–
	Mean(std)	309.7(22.9)	302.8(48.3)	431.9(102.3)	308.0(52.6)	–
	Coverage	0.790	1.000	0.750	1.000	–
$N_2 = 400$	Med	409.1	406.7	534.4	419.3	–
	Mean(std)	410.0(25.3)	407.5(39.5)	551.6(117.9)	422.6(54.9)	379.0(89.7)
	Coverage	0.750	0.930	0.810	1.000	0.850
$N_3 = 500$	Med	513.1	512.7	651.5	531.5	–
	Mean(std)	514.0(26.8)	513.3(35.5)	670.4(129.5)	535.8(62.0)	505.6(79.2)
	Coverage	0.780	0.870	0.760	1.000	0.930
$N_4 = 600$	Med	610.2	610.1	760.6	653.5	–
	Mean(std)	611.3(28.4)	611.2(34.7)	781.6(140.8)	661.2(79.2)	590.2(67.3)
	Coverage	0.820	0.880	0.760	1.000	0.910
$N_5 = 700$	Med	718.6	727.6	885.0	820.4	–
	Mean(std)	720.0(32.0)	731.0(43.3)	909.1(156.6)	842.6(133.4)	693.6(63.6)
	Coverage	0.800	0.850	0.780	0.930	0.920

bias when the true models are either Model $M_0^{(B)}$ or Model $M_t^{(B)}$. Moreover, the estimates for Model $M_t^{(B)}$ with noninformative priors have smaller biases than that of Jolly's method. Inferences obtained from Model $M_t^{(B)}$ are not sensitive to the priors selected. When the prior mean of P_j, $\gamma_1/(\gamma_1 + \gamma_2)$, is approximately equal to $\bar{P} = \sum P_j/t$, the estimates obtained from Model $M_{tb}^{(B)}$ have the smallest biases among others and coverage probabilities are above 80% when the true models are

Table 3
Simulation results for 100 runs and the true model is $M_b^{(B)}$; $(P_1, P_2, P_3, P_4, P_5) = (0.2, 0.2, 0.2, 0.2, 0.2)$, the prior is *Beta*(0.5,0.5)

$\delta = 1.5$		$M_0^{(B)}$	$M_t^{(B)}$	$M_b^{(B)}$	$M_{tb}^{(B)}$	Jolly
$N_1 = 300$	Med	206.6	177.2	876.2	368.8	–
	Mean(Std)	207.0(13.9)	177.8(64.6)	863.9(139.5)	408.2(220.9)	–
	Coverage	0.000	0.480	0.010	1.000	–
$N_2 = 400$	Med	297.2	296.7	1080.9	706.7	–
	Mean(Std)	297.8(15.4)	297.8(30.0)	1062.3(157.8)	719.8(250.0)	282.9(50.0)
	Coverage	0.000	0.150	0.020	0.900	0.350
$N_3 = 500$	Med	384.8	384.6	1242.5	1016.5	–
	Mean(Std)	385.3(16.2)	385.4(24.7)	1220.5(170.9)	1008.6(251.6)	374.5(41.5)
	Coverage	0.010	0.040	0.030	0.380	0.230
$N_4 = 600$	Med	474.1	473.1	1401.3	1321.8	–
	Mean(Std)	474.8(17.2)	473.8(23.1)	1374.2(180.1)	1293.3(231.9)	461.0(37.1)
	Coverage	0.000	0.000	0.020	0.090	0.120
$N_5 = 700$	Med	570.6	573.4	1588.2	1622.9	–
	Mean(Std)	571.3(18.6)	575.2(26.1)	1546.7(183.2)	1575.6(183.3)	559.2(36.3)
	Coverage	0.000	0.060	0.020	0.030	0.050

Table 4
Simulation results for 100 runs and the true model is $M_b^{(B)}$; $(P_1, P_2, P_3, P_4, P_5) = (0.2, 0.2, 0.2, 0.2, 0.2)$, the prior is *Beta*(10,40)

$\delta = 1.5$		$M_0^{(B)}$	$M_t^{(B)}$	$M_b^{(B)}$	$M_{tb}^{(B)}$	Jolly
$N_1 = 300$	Med	210.9	274.1	413.2	305.6	–
	Mean(std)	211.5(14.1)	273.1(35.2)	425.7(98.2)	308.3(52.8)	–
	Coverage	0.020	0.950	0.800	1.000	–
$N_2 = 400$	Med	299.4	341.0	533.0	421.0	–
	Mean(std)	299.8(15.5)	341.0(26.9)	547.6(112.9)	423.7(55.5)	285.8(51.2)
	Coverage	0.010	0.420	0.790	1.000	0.440
$N_3 = 500$	Med	390.6	409.1	646.3	532.5	–
	Mean(std)	391.2(16.5)	409.8(24.0)	662.9(123.7)	536.6(62.4)	367.7(39.4)
	Coverage	0.000	0.110	0.780	1.000	0.160
$N_4 = 600$	Med	485.5	497.6	760.9	653.8	–
	Mean(std)	486.0(17.2)	498.2(22.8)	779.5(134.4)	662.1(78.0)	469.9(38.1)
	Coverage	0.020	0.070	0.680	0.990	0.150
$N_5 = 700$	Med	571.1	582.8	872.2	813.1	–
	Mean(std)	571.8(18.2)	584.4(25.1)	891.8(148.7)	834.5(129.9)	552.2(34.9)
	Coverage	0.000	0.090	0.750	0.960	0.030

either Models $M_{tb}^{(B)}$ or $M_b^{(B)}$. Therefore, the estimator associated with Model $M_{tb}^{(B)}$ behaves nicely when we are able to have more information about the priors. When there exists a behavior response, the inferences obtained from Models $M_{tb}^{(B)}$ and $M_b^{(B)}$ have a sensitive dependence upon the hyper-parameters of the prior distribution of the capture probabilities. Hence, the inferences obtained from behavior response models ($M_{tb}^{(B)}$ or $M_b^{(B)}$) associated with the noninformative prior, i.e. $(\gamma_1, \gamma_2) = (0.5, 0.5)$,

Table 5
Simulation results for 100 runs and the true model is $M_b^{(B)}$; $(P_1, P_2, P_3, P_4, P_5) = (0.2, 0.2, 0.2, 0.2, 0.2)$, the prior is $Beta(0.5, 0.5)$

$\delta = 0.8$		$M_0^{(B)}$	$M_t^{(B)}$	$M_b^{(B)}$	$M_{tb}^{(B)}$	Jolly
$N_1 = 300$	Med	391.1	276.1	887.5	360.1	–
	Mean(Std)	392.3(32.0)	277.0(115.0)	874.0(139.4)	399.5(220.0)	–
	Coverage	0.250	0.990	0.010	1.000	–
$N_2 = 400$	Med	506.1	480.2	1084.2	690.2	–
	Mean(Std)	507.5(35.3)	480.9(60.7)	1066.8(156.9)	705.8(253.8)	486.6(142.2)
	Coverage	0.190	0.660	0.020	0.880	0.980
$N_3 = 500$	Med	617.7	607.7	1256.6	1004.1	–
	Mean(Std)	619.1(38.0)	608.7(51.3)	1234.2(169.4)	996.1(256.9)	595.0(109.7)
	Coverage	0.160	0.460	0.020	0.470	1.000
$N_4 = 600$	Med	724.3	724.4	1414.5	1317.4	–
	Mean(Std)	725.7(40.0)	726.7(50.3)	1387.8(178.2)	1287.5(234.8)	704.5(97.2)
	Coverage	0.170	0.250	0.010	0.080	0.970
$N_5 = 700$	Med	835.4	856.1	1589.0	1624.7	–
	Mean(Std)	837.6(45.2)	861.9(64.3)	1549.1(181.0)	1578.3(181.9)	788.1(86.1)
	Coverage	0.150	0.210	0.020	0.010	0.890

Table 6
Simulation results for 100 runs and the true model is $M_b^{(B)}$; $(P_1, P_2, P_3, P_4, P_5) = (0.2, 0.2, 0.2, 0.2, 0.2)$, the prior is $Beta(10, 40)$

$\delta = 0.8$		$M_0^{(B)}$	$M_t^{(B)}$	$M_b^{(B)}$	$M_{tb}^{(B)}$	Jolly
$N_1 = 300$	Med	381.8	317.1	418.3	306.8	–
	Mean(Std)	383.0(30.4)	319.1(55.9)	431.8(100.8)	309.6(52.7)	–
	Coverage	0.310	1.000	0.770	1.000	–
$N_2 = 400$	Med	491.4	450.7	532.4	420.5 –	
	Mean(Std)	492.4(32.8)	451.9(47.3)	548.6(115.2)	423.3(55.2)	480.7(140.5)
	Coverage	0.250	0.770	0.790	1.000	0.970
$N_3 = 500$	Med	601.2	571.3	647.7	532.0	–
	Mean(Std)	602.7(35.3)	572.6(43.6)	665.6(127.1)	536.5(62.8)	583.6(105.4)
	Coverage	0.190	0.540	0.780	1.000	1.000
$N_4 = 600$	Med	704.4	682.7	756.1	652.6	–
	Mean(Std)	705.7(37.9)	684.7(45.0)	775.9(138.4)	661.2(79.8)	669.2(88.4)
	Coverage	0.240	0.520	0.790	1.000	0.970
$N_5 = 700$	Med	824.0	833.2	882.2	822.1	–
	Mean(Std)	826.2(43.5)	838.3(59.9)	905.1(154.2)	845.3(135.7)	805.1(89.7)
	Coverage	0.220	0.250	0.710	0.970	0.910

behave poorly. Therefore, the noninformative priors are appropriate for estimating population size for Model $M_0^{(B)}$ and Model $M_t^{(B)}$. We do not recommend using the inferences obtained from Models $M_{tb}^{(B)}$ and $M_b^{(B)}$ associated with the noninformative priors when there exists a behavior response in the sampling experiment. When the prior mean of P_j is approximately equal to $\bar{P}$ and $\delta > 1$ (trap happy), Jolly's method has a larger bias and a lower coverage probability than that of the Bayesian approach

Table 7
Simulation results for 100 runs and the true model is $M_t^{(B)}$; $(P_1, P_2, P_3, P_4, P_5) = (0.3, 0.2, 0.1, 0.15, 0.25)$, the prior is $Beta(0.5, 0.5)$

$\delta = 1.0$		$M_0^{(B)}$	$M_t^{(B)}$	$M_b^{(B)}$	$M_{tb}^{(B)}$	Jolly
$N_1 = 300$	Med	383.2	239.8	225.3	403.3	–
	Mean(std)	383.6(18.2)	241.9(87.6)	226.0(44.0)	438.8(219.0)	–
	Coverage	0.030	0.960	0.190	0.990	–
$N_2 = 400$	Med	402.7	401.4	257.5	723.8	–
	Mean(std)	403.1(18.0)	402.3(41.3)	259.4(36.0)	732.8(245.6)	387.8(73.8)
	Coverage	0.700	0.880	0.060	0.810	0.860
$N_3 = 500$	Med	418.4	506.4	273.5	1008.8	–
	Mean(std)	419.1(18.9)	507.5(41.3)	276.6(34.1)	999.8(246.8)	487.1(96.6)
	Coverage	0.120	0.940	0.070	0.460	0.910
$N_4 = 600$	Med	521.9	603.7	364.8	1301.3	–
	Mean(std)	523.0(25.7)	605.1(40.7)	367.6(41.4)	1269.5(238.9)	587.7(85.0)
	Coverage	0.280	0.930	0.060	0.140	0.850
$N_5 = 700$	Med	850.6	724.4	567.4	1606.0	–
	Mean(std)	852.2(41.7)	727.6(45.3)	571.3(71.8)	1554.0(198.9)	690.0(60.5)
	Coverage	0.090	0.810	0.160	0.040	0.870

Table 8
Simulation results for 100 runs and the true model is $M_t^{(B)}$; $(P_1, P_2, P_3, P_4, P_5) = (0.3, 0.2, 0.1, 0.15, 0.25)$, the prior is $Beta(10, 40)$

$\delta = 1.0$		$M_0^{(B)}$	$M_t^{(B)}$	$M_b^{(B)}$	$M_{tb}^{(B)}$	Jolly
$N_1 = 300$	Med	386.3	359.6	307.0	348.2	–
	Mean(std)	386.9(17.6)	358.7(39.0)	310.4(36.9)	349.3(40.9)	–
	Coverage	0.040	0.870	0.990	0.940	–
$N_2 = 400$	Med	405.1	420.4	325.2	403.4	–
	Mean(std)	405.7(17.4)	420.8(32.1)	329.1(36.4)	406.2(41.4)	389.0(73.2)
	Coverage	0.710	0.920	0.490	1.000	0.890
$N_3 = 500$	Med	421.5	490.7	339.9	456.4	–
	Mean(std)	422.1(18.4)	491.7(34.4)	343.9(37.2)	460.7(50.0)	520.7(108.2)
	Coverage	0.170	0.930	0.160	1.000	0.960
$N_4 = 600$	Med	522.9	583.9	440.0	575.0	–
	Mean(std)	524.0(25.0)	585.6(38.3)	444.1(40.9)	586.0(82.6)	576.8(80.5)
	Coverage	0.280	0.860	0.160	1.000	0.870
$N_5 = 700$	Med	844.6	724.8	698.6	903.9	–
	Mean(std)	846.4(40.2)	727.7(43.1)	705.7(70.6)	931.7(168.9)	683.0(58.7)
	Coverage	0.140	0.760	0.900	0.650	0.880

(see Tables 4 and 10). In contrast, when the prior mean of P_j is approximately equal to $\bar{P}$ and $\delta < 1$ (trap shy), the estimate of population size for Jolly's method severely overestimates the true population size for each sampling period and with a high coverage probability due to the large variations (see Tables 6 and 12).

Table 9
Simulation results for 100 runs and the true model is $M_{tb}^{(B)}$; $(P_1, P_2, P_3, P_4, P_5) = (0.3, 0.2, 0.1, 0.15, 0.25)$, the prior is $Beta(0.5, 0.5)$

$\delta = 1.5$		$M_0^{(B)}$	$M_t^{(B)}$	$M_b^{(B)}$	$M_{tb}^{(B)}$	Jolly
$N_1 = 300$	Med	278.9	192.8	224.9	391.5	–
	Mean(std)	279.0(11.3)	194.0(58.5)	227.8(42.5)	430.2(222.1)	–
	Coverage	0.480	0.530	0.290	1.000	–
$N_2 = 400$	Med	297.1	302.3	253.4	717.0	–
	Mean(std)	297.3(10.6)	303.5(26.5)	257.5(35.5)	726.8(252.0)	286.8(39.9)
	Coverage	0.000	0.200	0.070	0.870	0.300
$N_3 = 500$	Med	311.5	398.5	270.6	1014.2	–
	Mean(std)	311.8(11.1)	398.7(27.1)	275.8(33.5)	1001.2(247.5)	396.8(62.2)
	Coverage	0.000	0.050	0.050	0.450	0.550
$N_4 = 600$	Med	406.5	466.0	359.6	1305.7	–
	Mean(std)	407.1(15.1)	466.5(25.4)	364.5(40.3)	1277.3(230.8)	448.8(46.8)
	Coverage	0.000	0.000	0.030	0.080	0.140
$N_5 = 700$	Med	656.6	553.3	566.8	1610.8	–
	Mean(std)	657.5(24.4)	554.7(23.4)	575.0(71.5)	1564.2(187.2)	540.4(32.4)
	Coverage	0.520	0.010	0.290	0.000	0.010

Table 10
Simulation results for 100 runs and the true model is $M_{tb}^{(B)}$; $(P_1, P_2, P_3, P_4, P_5) = (0.3, 0.2, 0.1, 0.15, 0.25)$, the prior is $Beta(10, 40)$

$\delta = 1.5$		$M_0^{(B)}$	$M_t^{(B)}$	$M_b^{(B)}$	$M_{tb}^{(B)}$	Jolly
$N_1 = 300$	Med	288.0	312.1	311.6	346.8	–
	Mean(std)	288.1(11.5)	311.2(27.1)	315.0(38.6)	347.8(40.6)	–
	Coverage	0.570	0.970	0.970	0.940	–
$N_2 = 400$	Med	307.0	350.6	330.1	402.5	–
	Mean(std)	307.2(10.7)	351.0(23.0)	334.3(38.3)	405.1(41.1)	291.8(40.4)
	Coverage	0.020	0.420	0.520	1.000	0.260
$N_3 = 500$	Med	321.2	406.0	344.9	456.1	–
	Mean(std)	321.5(11.3)	406.8(24.5)	349.4(39.2)	460.1(49.8)	396.3(60.4)
	Coverage	0.000	0.060	0.160	1.000	0.500
$N_4 = 600$	Med	421.0	482.2	447.8	577.1	–
	Mean(std)	421.6(15.6)	483.1(26.3)	452.6(43.7)	588.6(83.9)	457.9(47.5)
	Coverage	0.000	0.040	0.240	1.000	0.220
$N_5 = 700$	Med	669.9	587.9	716.4	913.7	–
	Mean(std)	670.9(24.8)	589.4(25.7)	724.7(75.8)	942.1(172.4)	549.7(33.1)
	Coverage	0.630	0.110	0.930	0.670	0.010

3.2. Real example: HIV serosurveillance data

In this subsection, we estimate the numbers of intravenous drug users (IVDU's) in Thailand infected with human immunodeficiency virus (HIV) who have not progressed to AIDS from the semi-annual HIV serosurveillance data of Thailand from June 1991 to June 1993. The nation-wide serosurveillance date is published by

Table 11
Simulation results for 100 runs and the true model is $M_{tb}^{(B)}$; $(P_1, P_2, P_3, P_4, P_5) = (0.3, 0.2, 0.1, 0.15, 0.25)$, the prior is *Beta*(0.5,0.5)

$\delta = 0.8$		$M_0^{(B)}$	$M_t^{(B)}$	$M_b^{(B)}$	$M_{tb}^{(B)}$	Jolly
$N_1 = 300$	Med	470.0	292.6	238.2	378.3	–
	Mean(std)	470.7(25.0)	293.2(111.5)	240.4(45.6)	418.9(222.1)	–
	Coverage	0.000	1.000	0.290	1.000	–
$N_2 = 400$	Med	492.8	486.9	270.6	709.2	–
	Mean(std)	493.6(25.2)	487.8(56.6)	274.5(38.0)	720.2(251.8)	467.8(105.9)
	Coverage	0.070	0.690	0.140	0.890	1.000
$N_3 = 500$	Med	512.5	614.8	287.6	1006.0	–
	Mean(std)	513.1(26.4)	615.9(53.9)	292.5(36.4)	995.4(251.0)	616.8(152.0)
	Coverage	0.780	0.460	0.080	0.430	1.000
$N_4 = 600$	Med	619.6	721.5	379.9	1302.8	–
	Mean(std)	621.4(34.6)	723.6(54.2)	384.7(43.5)	1271.3(237.5)	695.8(119.3)
	Coverage	0.730	0.380	0.070	0.090	0.990
$N_5 = 700$	Med	1019.1	859.2	588.4	1610.2	–
	Mean(std)	1022.1(58.2)	863.8(63.3)	595.1(73.9)	1561.0(194.7)	800.6(83.7)
	Coverage	0.000	0.240	0.280	0.030	0.860

Table 12
Simulation results for 100 runs and the true model is $M_{tb}^{(B)}$; $(P_1, P_2, P_3, P_4, P_5) = (0.3, 0.2, 0.1, 0.15, 0.25)$, the prior is *Beta*(10,40)

$\delta = 0.8$		$M_0^{(B)}$	$M_t^{(B)}$	$M_b^{(B)}$	$M_{tb}^{(B)}$	Jolly
$N_1 = 300$	Med	457.0	378.4	311.8	340.5	–
	Mean(std)	457.7(23.7)	377.4(48.0)	315.7(40.8)	341.5(40.3)	–
	Coverage	0.000	0.830	0.990	0.990	–
$N_2 = 400$	Med	479.3	460.1	331.4	397.0	–
	Mean(std)	480.0(23.7)	460.9(39.1)	336.0(40.4)	399.5(40.6)	478.9(113.0)
	Coverage	0.130	0.680	0.520	1.000	0.990
$N_3 = 500$	Med	498.4	543.2	346.4	450.4	–
	Mean(std)	499.1(24.8)	544.1(43.0)	351.4(41.5)	454.4(49.6)	595.7(143.6)
	Coverage	0.720	0.870	0.200	0.990	1.000
$N_4 = 600$	Med	604.9	658.9	446.6	570.2	–
	Mean(std)	606.2(33.0)	661.4(50.5)	452.2(45.8)	581.2(82.5)	669.4(111.7)
	Coverage	0.790	0.770	0.220	1.000	0.980
$N_5 = 700$	Med	1002.6	841.5	724.9	910.1	–
	Mean(std)	1005.0(55.3)	846.2(61.1)	734.5(81.1)	938.4(173.6)	805.0(85.2)
	Coverage	0.000	0.240	0.980	0.690	0.870

Division of Epidemiology, Ministry of Public Health (MOPH) of Thailand. The data for IVDU's used in this example is listed in Table 13.

The population size to be estimated is the number of HIV-infecteds within the IVDU's population. We assume no natural deaths between the samples, as the population of IVDU's consists of young adults with low mortalities. Moreover, the sampling period is two and half years. Due to the long incubation period of HIV, it

Table 13
The sampling period and HIV-infected ratio

Sampling period		June 91	December 91	June 92	December 92	June 93
IVDU's sample	e_i	2933	2383	2668	2686	3515
HIV-infected	u_i	931	912	952	1039	1234
Ratio (%)		38.41	31.74	38.27	38.68	35.11

is reasonable that we can assume no deaths during the sampling period. The births only model with no recapture is the model to work with. For more detail, those who test positive will not be tested again, so that there is no recapture and we have a special case of the births only model. During ith sampling period, $i = 1, 2, \ldots, 5$, let the IVDU's sample and HIV-infected be denoted e_i and u_i, respectively. $u_1, \ldots, u_5$ are available rather than $m_1, \ldots, m_5$. We can estimate the number of HIV-infected IVDU's by using either Model $M_b^{(B)}$ or Model $M_{tb}^{(B)}$. In the simulation study, it shows that if we have information about the prior distributions then the estimators associated with Model $M_{tb}^{(B)}$ perform better than $M_b^{(B)}$. Moreover, the observed "capture probability" in each sampling period is not the same. We decide to estimate the population size of HIV-infected by using Model $M_{tb}^{(B)}$.

In the following procedure, we will utilize the prior information, P_j, from Mastro et al. (1994) and Weniger et al. (1991). Mastro et al. (1994) estimated the number of HIV-infected IVDU's in Bangkok to be approximately 12,000 by using two sets of data on IVDU's in Bangkok to obtain an estimated number of IVDU's in Bangkok of 32,574. Weniger et al. (1991) reported that in fiscal year 1989 there were 60,323 admissions for treatment at 138 registered heroin/opiate detoxification centers in Thailand, out of which 27,056 admissions are in Bangkok. Assuming that the number of IVDU's in Bangkok maintain roughly the same proportion when compared with the nation-wide total in 1991, we obtain an estimate of IVDU's in Thailand of 72,626. Suppose that 72,626 is also the nation-wide total in June 1991. As all of the ratios of u_i and e_i are approximately constant, we can assume that the prior of the capture probability for the IVDU's population is the same as the prior of the capture probability for HIV-infected within the IVDU's population. Therefore, the prior of P_j can be obtained from the information of e_i and the nation-wide total in June 1991 for the number of IVDU's. If P_j follows $Beta(\gamma_1, \gamma_2)$, then its mean and coefficient of variation are $\gamma_1/(\gamma_1 + \gamma_2)$ and $\sqrt{\gamma_2/(\gamma_1(\gamma_1 + \gamma_2 + 1))}$, respectively. The estimates of $\gamma_1/(\gamma_1 + \gamma_2)$ and $\sqrt{\gamma_2/(\gamma_1(\gamma_1 + \gamma_2 + 1))}$ are

$$\frac{\gamma_1}{(\gamma_1 + \gamma_2)} \cong 0.04 \left(= \frac{e_1}{72626} \right) \tag{17}$$

and

$$\frac{\gamma_2}{\gamma_1(\gamma_1 + \gamma_2 + 1)} = \frac{\sum_{j=1}^{5} (e_j - \bar{e})^2 / 5}{\bar{e}^2} \cong 0.018, \tag{18}$$

where $\bar{e}$ is the mean of e_i. We can solve (γ_1, γ_2) from equations (17) and(18) which are (53.29,1279.04). Notice that in the Gibbs sampler we divided the numbers by 20

Table 14
Bayesian estimates of the number of HIV-infected IVDU's in Thailand

Date	Median	Mean	Std	95% Bayesian interval
June 1991	19,831	19,726	1325	(16,917–22,333)
December 1991	22,884	22,837	1048	(20,794–24,784)
June 1992	25,758	25,808	1097	(23,632–27,736)
December 1992	28,911	28,873	1073	(26,911–30,843)
June 1993	32,171	32,140	845	(30,430–33,552)

to draw the Gibbs sequence from the conditional posteriors. The posterior median, posterior mean, and 95% Bayesian intervals are listed in Table 14. These Bayes estimates are based on Monte Carlo samples from the Gibbs sampler run of 5000 iterations after 3500 burn-in, and selecting every 5th sampled value.

It is assumed that no natural deaths occurred in the births only model. This assumption is plausible when the time span of the samples is short. For a comparison, we applied the same proportion of IVDU's in Bangkok (Weniger et al., 1991) to obtain the estimates of HIV-infected IVDV's in Bangkok. We obtain an estimate of $10,264$ HIV-infected IVDU's in Bangkok by using the posterior median in December 1991 which is slightly less than $12,000$ obtained by Mastro et al. (1994).

4. Concluding remarks

In this paper, we propose a unified approach for estimating the population size at each sampling period for Models $M_0^{(B)}$, $M_t^{(B)}$, $M_b^{(B)}$, and $M_{tb}^{(B)}$. The population size at the first sampling period N_1 is unestimable using traditional maximum likelihood estimation for Model $M_t^{(B)}$. There also exists an unidentifiability problem if the information of the first capture is employed to estimate the population size for Model $M_{tb}^{(B)}$. The Bayesian approach enables us to estimate more parameters than observations at hand (see Lee and Chen, 1998). Therefore, the unidentifiability problem can be resolved for Model $M_{tb}^{(B)}$ using the proposed Bayesian approach. The main advantage of using the proposed procedure is that we are able to estimate N_1 via prior information for the four models. The results obtained in simulation study shows that the Gibbs sampler offers reasonable inferences for population size in general. Inferences obtained from behavior response models such as Model $M_b^{(B)}$ or Model $M_{tb}^{(B)}$ have a sensitive dependence on the hyper-parameters of the prior distribution of the capture probabilities. However, the inference performs satisfactorily when we make an appropriate choice of hyper-parameters of the prior distribution.

Model selection is an important issue in the estimation of population size which is beyond the scope of this paper and shall be considered in subsequent work. The results of this paper can be used to estimate the population size on a closed population under the simplifying condition $B_1 = \cdots = B_{s-1} = 0$. Hsieh et al. (1999) further extends

the proposed method to a generalized removal model which allows for recruitment and deaths to occur during the experiment. Our inference for Model $M_t^{(B)}$ generalizes the results of George and Robert (1992). Moreover, the inference for Model $M_b^{(B)}$ and Model $M_{tb}^{(B)}$ generalizes the results of Lee and Chen (1998). Finally, the results of the proposed approach also can be generalized to analyze censor or truncated data sets.

Acknowledgements

C.W.S. Chen and S.-M. Lee acknowledge research support from National Science Council of Taiwan grants NSC86-2115-M-035-017 and NSC87-2118-M-035-001, respectively.

References

Casella, G., George, E.I., 1992. Explaining the Gibbs sampler. Amer. Statist. 46, 167–174.

Cormack, R.M., 1968. The statistics of capture–recapture methods. Oceanogr. Mar. Biol. Ann. Rev. 6, 455–506.

Cormack, R.M., 1972. The logic of capture–recapture estimates. Biometrics 28, 337–343.

Darroch, J.N., 1959. The multiple recapture census: II. Estimation when there is immigration or death. Biometrika 46, 336–352.

Gelfand, A.E., Smith, A.F.M., 1990. Sampling-based approach to calculating marginal densities. J. Amer. Statist. Assoc. 85, 398–409.

Gelfand, A.E., Hills, S.E., Racine-Poon, A., Smith, A.F.M., 1990. Illustration of Bayesian inference in normal data models using Gibbs sampling. J. Amer. Statist. Assoc. 85, 972–985.

Geman, S., Geman, D., 1994. Stochastic relaxation, Gibbs distributions and the Bayesian restoration of images. IEEE Trans. Pat. Anal. Mach. Intell. 6, 721–741.

George, E.I., Robert, C.P., 1992. Capture–recapture estimation via Gibbs sampling. Biometrika 79, 677–683.

Hsieh, Y.-H., Chen, C.W.S., Lee, S.-M., 1999. Empirical Bayes approach to estimating the number of HIV-infected individuals in hidden and elusive populations. Tech. Rep. (submitted)

Hwang, W.-D., Chao, A., 1995. Quantifying the effects of unequal catchabilities on Jolly–Seber estimators via sample coverage. Biometrics 51, 128–141.

Jolly, G.M., 1965. Explicit estimates from capture–recapture data with both death and immigration-stochastics model. Biometrika 52, 225–247.

Lee, S.-M., Chen, C.W.S., 1998. Bayesian inference of population size for behavioral response models. Statist. Sinica 8, 1233–1247.

Mastro, T.D., Kitayaporn, D., Weniger, B.G., Vanichseni, S., Laosunthorn, V., Uneklabh, T., Uneklabh, C., Choopaya, K., Limpakarnjanarat, K., 1994. Estimating the number of HIV-infected injection drug users in Bangkok: A capture–recapture method. Amer. J. Public Health 84, 1094–1099.

Pollock, K.H., Nichols, J.D., Brownie, C., Hines, J.E., 1990. Statistical inference for capture–recapture experiments. Wildlife Monographs 107, 1–97.

Seber, G.A.F., 1982. The Estimation of Animal Abundance, 2nd ed. Griffin, London.

Seber, G.A.F., 1986. A review of estimating animal abundance. Biometrics 42, 267–292.

Seber, G.A.F., 1992. A review of estimating animal abundance II. Internat. Statist. Rev. 60, 129–166.

Tanner, M., 1994. Tools for Statistical Inference: Methods for the Exploration of Posterior Distributions and Likelihood Functions. Springer, New York.

Tanner, M.A., Wong, W.H., 1987. The calculation of posterior distributions by data augmentation. (with discussion). J. Amer. Statist. Assoc. 82, 528–550.

Tierney, L., 1994. Markov chains for exploring posterior distributions. Ann. Statist. 22, 1701–1762.
Weniger, B., Limpakarnjanarak, K., Ungchusak, K., Thanprasertsuk, S., Choopanya, K., Vanichseni, S., Thongcharoen, P., Wasi, C., 1991. The epidemiology of HIV infection and AIDS in Thailand. AIDS 5 (Suppl. 2), s71–s85.